Conventional and Robust Norming in identifying Preclinical Dementia

Ellen Grober[1], **Wenzhu Mowrey**[2], **Mindy Katz**[1], **Carol Derby**[1], **Richard B. Lipton**[1,2]

[1]Department of Neurology and the Einstein Aging Study, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, New York

[2]Department of Epidemiology and Population Health, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, New York

Abstract

Objective: To contrast four approaches to norming two widely used memory tests in older adults for purposes of detecting preclinical dementia.

Methods: The study sample included participants from the Einstein Aging Study who were over age 70, free of dementia at baseline and followed for at least 5 years. Norms were derived from a conventional sample (excluding individuals with dementia at baseline but not those who developed dementia during follow-up) and a robust normative sample (excluding persons with dementia at baseline as well as those who developed dementia over 5 years of follow-up). Both normative samples were examined with and without adjustment for age and education. We contrasted the picture version of the Free and Cued Selective Reminding Test with Immediate recall (pFCSRT +IR) and the Logical Memory Test (LM) test for their ability to identify persons with preclinical dementia, operationally defined by the development of diagnosable dementia over 5 years of follow-up, using these four approaches to developing norms for detecting preclinical dementia.

Results.—Of 418 participants included in the conventional normative sample the mean age was 78.2 and 59% were female. There were 78 incident cases of dementia over 5 years leaving 340 participants in the robust normative sample. Means and SDs were defined for both the conventional and robust normative samples and cut-scores with and without adjustment were set at 1.5 SD below the mean of each test. As predicted, in comparison with the conventional sample, the robust sample had higher cut-scores, which provided higher sensitivity for detecting preclinical dementia. This effect persisted regardless of adjustment. The pFCSRT+IR was more sensitive than LM in detecting incident dementia cases.

Conclusion: When using cognitive test norms to identify preclinical dementia, robust norming procedures improves detection using both the FCSRT and LM.

Ellen Grober, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, Phone/fax: 914-963-5602, ellen.grober@einstein.yu.edu.

Keywords

MCI; memory; incident dementia; robust and conventional norms; Logical Memory; Free and Cued Selective Reminding Test; preclinical dementia

Introduction

The number of people with AD dementia is expected to increase dramatically over the next 40 years (Hebert, Weuve, Scherr, & Evans, 2013), continuing a trend that reflects the aging of the US population (Brookmeyer, 2007). In anticipation of this trend, various criteria have been proposed for identifying individuals at increased risk of future dementia and AD. These criteria usually rely on memory complaints and objective evidence of memory impairment with preserved everyday functioning (amnestic mild cognitive impairment or aMCI: Petersen et al., 1999), sometimes combined with neuroimaging or biomarker evidence (Dubois et al, 2007; Albert et al, 2011).

Various strategies have been used to norm memory tests. "Conventional" norms are developed based on the distribution of cognitive scores in dementia-free samples. This popular approach, sometimes termed comparative norming, may be appropriate for comparing individuals to their peers. However, although conventional normative samples exclude prevalent dementia cases, they include individuals who subsequently develop dementia on follow-up. At cross-section, individuals who develop dementia over the next 5 to 7 years, on average, have reduced performance on cognitive tests relative to their peers. Thus, including them in the normative sample underestimates the level of performance and overestimates the variance of cognitive tests (Sliwinski, Lipton, Buschke, & Stewart, 1996).

To mitigate this problem, we and others have advocated the development of robust norms in longitudinal samples from which individuals who develop dementia over several years of follow-up are removed (Sliwinski et al, 1996; _S1_Reference15Ivnik, Smith, & Lucas, 1997; De Santi et al, 2008; Holtzer et al., 2008). Individuals who remained dementia-free for at least four years performed significantly better on neuropsychological tests at baseline than individuals who developed dementia during the same period (De Santi et al, 2008; Holtzer et al, 2008). In the De Santi study, robust norms were more sensitive than conventional norms at identifying those healthy individuals who declined to MCI or dementia. In a study by Ritchie et al (2007), conventional norms underestimated performance on all 12 neuropsychological measures and overestimated test variance in 7 of the 12 compared with robust norms that excluded incident dementia cases that developed over 10 years. Because of the overlapping samples, between-group comparisons were not undertaken (Ritchie et al, 2007).

A second factor that affects the derivation of norms for identifying incident dementia is the adjustment of test score performance for the influence of age and education. Though adjustments for age and education seem appropriate in the context of comparing individuals to their peers, statistically removing the contribution of these dementia risk factors from memory test scores can severely reduce discriminative validity for future onset of AD (Sliwinski, Buschke, Stewart, Masur, & Lipton, 1997). As shown in Sliwinski et al., age

adjusted memory test scores had a sensitivity for dementia that was 28% lower than unadjusted scores.

Herein, we compare four strategies for identifying incident dementia using two memory tests that have been shown previously to identify persons at risk for future dementia (Grober et al, 2000; Sarazin et al, 2007; Derby et al, 2014; Wagner et al, 2012). We compare prediction using conventional norms with and without age and education adjustments to robust norms with and without adjustments. The tests are the picture version of the Free and Cued Selective Reminding Test with Immediate Recall (pFCSRT+IR) and the Logical Memory (LM) subtest of the Wechsler Memory Scale – Revised. The purpose of these comparisons is to determine the optimal strategy for identifying persons with pre-dementia, operationally defined here as the development of diagnosable dementia over 5 years of follow-up.

In previous comparisons, pFCSRT+IR outperformed delayed story recall in predicting the CSF AD profile among MCI patients (Wagner et al, 2012) and outperformed immediate story recall in predicting AD during 3 years of follow-up among EAS patients with memory complaints (Derby et al, 2013). In neither study, did the addition of LM improve the prediction of future AD over pFCSRT+IR alone. We predict that pFCSRT+IR will outperform LM in detecting incident dementia over five years of follow-up when matched for the high specificity needed in primary care settings. Both LM Longitudinal data from the Einstein Aging Study (EAS), a community based cohort, was used to test the following hypotheses: 1) estimates of mean cognitive test performance will be higher and between person variance will be lower in the robust sample in comparison to the conventional sample, resulting in higher cut scores; 2) the robust norms will have higher sensitivity for identifying incident dementia than conventional norms; and 3) unadjusted cut-scores will perform better than cut-scores adjusted for age and education in identifying incident dementia.

Methods

Study Population.

Longitudinal data collected from participants in the Einstein Aging Study (EAS), a systematically recruited cohort of adults from a multi-ethnic, community-dwelling population in Bronx County, NY, provided the bases for these analyses. Detailed study methods have been described previously (Katz et al, 2012). The cohort includes individuals who are at least 70 years of age, Bronx residents, non-institutionalized, and English speaking and provided written consent according to protocols approved by Einstein's institutional review board. In-person neuropsychological evaluations completed at baseline and annually include pFCSRT+IR and LM performance. Assessments also include a standardized neurological evaluation, demographic information, medical history, medication use, health behavior, instrumental activities of daily living (ADLs), and informant interviews whenever possible.

The present analyses were restricted to individuals who were dementia-free at baseline and who had at least five years of follow-up or who developed dementia within five years. A

"conventional normative sample" was defined as all individuals meeting these criteria. A "robust" normative sample was defined by excluding individuals from the conventional sample who developed clinical dementia within five years of follow-up. The five-year window was chosen because of the high predictive validity of the pFCSRT+IR in this pre-dementia period (Grober et al, 2000; Grober et al, 2010; Derby et al, 2012).

Dementia Diagnosis.

A diagnosis of dementia was based on standardized clinical criteria from the Diagnostic and Statistical Manual, Fourth Edition (DSM-IV: American Psychiatric Association, 1994) and required impairment in memory plus at least one additional cognitive domain, accompanied by evidence of functional decline. Diagnoses were assigned at consensus case conferences that included comprehensive review of all cognitive test results, relevant neurological signs and symptoms, informant responses, and functional status by two neurologists and a neuropsychologist.

Memory tests.

The pFCSRT+IR (Grober, Buschke, Crystal, Bang, & Dresner, 1988; Grober, Lipton, Hall, & Crystal, 2000) begins with a study phase in which participants search a card containing four pictures (e.g., grapes) for an item that goes with a unique category cue (e.g., fruit). After all four items are identified, the card is removed and immediate cued recall of the four items is tested while the items are still in working memory. The search continues for the next group of four items until all 16 items have been identified and retrieved in immediate recall. The test phase includes three recall trials, each consisting of free recall followed by cued recall for items not retrieved by free recall for a maximum score of 48. Each separate trial is followed by 20 seconds of interference to purge working memory. Controlled learning in the study phase ensures attention, promotes deep semantic processing, and maximizes recall in the test phase through encoding specificity. The dependent measure in these analyses is free recall summed over the three test trials, with higher score indicating better performance.

The LM subtest of the Wechsler Memory Scale – Revised (Wechsler, 1987) is comprised of two stories, each consisting of 25 elements. Each story is read to the participant who recalls the story elements immediately after hearing them. The dependent measure is the combined number of story elements recalled, with higher scores indicating better recall. Immediate recall was used because administration of delayed recall did not begin until 2008, more than 15 years after EAS data collection began.

Statistical analyses.

Demographics and memory test performance at baseline were summarized for the conventional, robust and incident dementia samples. Two sample t-tests, Pearson's Chi-square tests or Fisher's exact tests were used to compare the robust sample against the incident dementia samples. Cut-scores on the pFCSRT+IR and LM tests were computed from the conventional and robust normative samples based on the accepted practice of defining cognitive impairment as being 1.5 SD units below the mean performance of the normative sample (Petersen et al, 1999). Linear regression was used to compute age and education adjusted cut-scores (<12 versus >=12) for LM and pFCSRT+IR. For an individual

subject with a particular age and education (level ≥ 12 years or < 12 years), performance was estimated for this age and education based on linear regression parameters. The mean and standard deviation estimates were used to compute the age and education specific cutoff for the subject which were then compared to the raw score. The classification accuracy for identifying incident dementia under the four strategies was compared: unadjusted and adjusted cut scores based on the conventional normative sample and unadjusted and adjusted cut scores based on the robust normative sample. Sensitivity of pFCSRT+IR and LM to incident dementia at equivalent levels of specificity was evaluated with the McNemar's (1947) test. The statistical software SAS 9.3 (SAS Institute Inc., Cary, NC) was used for the analysis.

Results

The conventional normative sample included 418 participants who were dementia-free at baseline and had at least 5 years of follow-up. Of these 418 participants, 78 participants developed dementia within 5 years of baseline, comprising the incident dementia group. The 340 who remained dementia-free for at least 5 years comprised the robust normative sample. Table 1 shows the baseline characteristics by subgroup. Compared with the robust normal group, those who developed incident dementia were older (p<0.0001) and performed worse on both pFCSRT+IR (p<0.0001) and LM tests (p<0.0001) at baseline. As anticipated, scores were higher in the robust sample compared to the conventional sample for both tests and variance was lower. The difference in scores between the robust and conventional samples for 80 –89 year olds on both pFCSRT+IR (2.4 points (30.8–28.4)) and Logical Memory (1.9 points (19.4–17.5)) was double the difference for 70 to 79 year olds on pFCSRT+IR (1.1 points (32.5–31.2)) and on LM (0.8 points (21.8–21.0)). Because of the overlapping samples, between-group comparisons were not undertaken.

Free recall declined with age in the conventional and robust samples (r = −0.27, p<0.0001 versus r = −0.16, p<0.004) as did story recall (r = −.26 p<0.0001 versus r = −.17, p<0.001). Higher education enhanced free recall in the conventional and robust samples (r = 0.10, p<0.04 versus r = −.15, p=0.007) as well as in story recall (r = 0.37, p<0.0001 versus r= 0.41, p<0.0001).

Table 2 shows the linear regression equation for computing age and education adjusted cut-scores (<12 versus >=12) for LM and pFCSRT+IR from the conventional and robust samples. The unadjusted and adjusted cut scores were applied to pFCSRT+IR and LM raw scores to determine the classification accuracy of the four approaches in identifying the incident dementia cases. The classification results are shown in Table 3 along with the unadjusted cut-scores. Age and education adjusted cut-scores are plotted in Figure 1. The cut score derived from the robust normative sample was nearly 4 points higher (24.3 (5.1) versus 20.6 (6.4)) on pFCSRT+IR and nearly 2 points higher on LM (10.3 (7.2) versus 8.6 (7.5)) compared to the cut scores derived from the conventional normative sample. At high levels of specificity (>92%), sensitivity for the pFCSRT+IR was highest using the unadjusted cut-scores derived from the robust normative sample compared to unadjusted cut-scores derived from the conventional normative sample (57.7% versus 34.6%). Similarly, sensitivity for the LM was highest using unadjusted cut-scores derived from the robust normative sample

compared to unadjusted cut-scores derived from the conventional normative sample (32.0% versus 19.2%). Sensitivity was higher using unadjusted cut-scores compared to age and education-adjusted cut-scores for both tests in the robust sample (5.9% and 6.4%) and the conventional sample (1.2% and 2.6%).

Next we compared the sensitivities of pFSCRT+IR and LM using the strategy of unadjusted cut scores from a robust normative sample, which yielded the highest sensitivity for both tests. Test agreement with clinical diagnosis is shown in the 2X2 contingency table (Table 4) Of the 78 dementia cases, both tests correctly identified 17 and misidentified another 25. When the classifications of the two tests diverged, the pFCSRT+IR identified 28 individuals as incident cases of dementia that LM misclassified as not demented compared to 8 incident cases correctly classified by LM that were misclassified by the pFCSRT+IR. In other words, when classification differed, pFCSRT+IR identified 3½ times as many incident dementia cases as LM at equivalent high levels of specificity (>90%, (McNemar's test, statistic =11.1, p=.001).

Discussion

This paper has implications for characterizing normal aging and for the identification of preclinical dementia. Four different strategies were used to norm two widely used tests of episodic verbal memory. Norms were developed in a community cohort of individuals free of dementia at baseline and followed for at least 5 years. Cut scores defining cognitive impairment as being 1.5 SD units below the group mean performance were calculated from conventional and robust normative samples with and without adjustment for age and education. Our findings confirm that conventional norms mischaracterize normative aging; because persons with preclinical dementia are included, the approach underestimates cognitive performance in older adults (Sliwinski et al, 1996; De Santi et al, 2008; Holtzer et al, 2008). The inclusion of individuals with and without preclinical dementia in these normative samples results in an overestimation of variability in cognitive performance. Creation of robust normative samples by removal of persons who develop dementia during follow-up results in higher levels of cognitive performance and a reduction in variance. The effect of robust norming is substantially greater in 80 to 89 year olds than in 70 to 79 year olds, presumably because the incidence of dementia and preclinical dementia increases exponentially with age in persons over 65 (Brookmeyer, 2007).

The cut scores for identifying future incident dementia cases based on robust norms were 4 points higher for pFCSRT+IR and 2 points higher for LM than conventional norms. This change in cut-score had dramatic effects on sensitivities for preclinical dementia, as defined by the onset of incident dementia within 5 years. Sensitivity was higher using robust norms than conventional norms regardless of adjustment for age and education. The improved sensitivity of robust norms for both LM and pFCSRT+IR extends the findings described in De Santi et al (2008). De Santi showed that a composite score based on delayed recall of paragraphs, paired associates, and word lists from a robust sample was more sensitive at identifying decliners than the composite score from the conventional sample (32% versus 15%). The same or higher sensitivities were achieved in the current study using robust

norms for each test alone (LM: 32% and pFCSRT+IR: 58%) while maintaining high specificity (>90%) as in De Santi et al, 2008.

We expected that age and education adjustments would reduce sensitivities, because advanced age and low education are powerful risk factors for dementia (Sliwinski et al, 1997). Adjusting for these covariates removes their predictive power. Our data support this hypothesis but the magnitude of the effect is not as great as we expected based on prior work (Sliwinski et al, 1997). One reason why age and education adjustments reduced sensitivities only slightly may be that the current cohort is two years younger than the Sliwinski cohort. Another reason may be that different memory tests were used which, in turn, may have different associations with age.

Finally, pFCSRT+IR out-performed LM in identifying incident dementia cases. At the high levels of specificity needed in primary care, pFCSRT+IR identified 3½ times as many incident dementia cases as LM among those cases on which they disagreed.

Some of the incident dementia cases likely had mild cognitive impairment (MCI) when they entered the study. In a longitudinal study, persons who develop diagnosable Alzheimer's disease over five years probably had preclinical dementia at enrollment. Our goal is not to remove people with preclinical dementia from longitudinal samples but to better identify them prior to dementia diagnosis. We decided against removing baseline aMCI for several reasons. First, aMCI is an unstable diagnosis in population studies. Up to 40% of persons who develop incident aMCI at one study assessment remit by the next (Kaduszkiewicz et al, 2014). Eliminating MCI participants from the normative sample is likely to have two consequences: fewer incident cases and easier discrimination of incident cases from those with no memory impairment. Instead we chose to remove only incident dementia cases because their trajectory was certain.

Conventional and robust norms serve different purposes (Sliwinski et al, 1997). Conventional norms, which typically exclude persons who are in poor health or have prevalent dementia, are used to answer the question of how the performance of an individual compares to his or her age and education-matched peers. Robust norms endeavor to characterize cognitive trajectories in the absence of preclinical dementia and are useful for identifying persons at high risk of future dementia.

Optimal clinical cut scores depend on the nature and timing of the outcome of interest and the composition of the sample. In a longitudinal aging context, outcomes include aMCI, incident dementia, prodromal AD, all-cause dementia, AD, and nonAD dementia. The timing of the outcome could be the identification of current status or the prediction of future status. When predicting future status, the time horizon or prediction window will influence the selection of optimal cut scores (Derby et al, 2013). Using a cut score of <=24 on the pFCSRT+IR to predict future dementia, sensitivity declined from >85% at 5 years to 70% 3 years later (Grober et al, 2000). Specificity increased gradually from 80% at 5 years to 90% at 9 years. The cut score for samples enriched with persons at high risk for dementia will be lower than for robust normative samples because the discrimination will be easier to make. Cut scores for the pFCSRT+IR have been published for predicting future dementia and AD

(Grober et al, 2000; Derby et al, 2013) and for identifying prevalent dementia in both population and primary care cohorts (Grober et al, 1988; Grober et al, 2010; Grober et al, 2014).

Variations in test administration also influence optimal cut scores. There are four versions of the FCSRT. All use controlled learning to ensure attention and deep semantic processing and to maximize recall through encoding specificity. The major distinction is the form in which the to-be-remembered items are presented. The original version presents the items as simple line drawings (Grober & Buschke, 1987) whereas they are presented as printed words in the modified version (Sarazin et al, 2007). The other variation has to do with whether or not the study phase includes immediate cued recall. Immediate recall confirms correct initial encoding, demonstrates that the participant understands the task, and provides retrieval practice before the test phase (Grober & Buschke, 1987). It increases the opportunity that each item has an adequate chance to be encoded into the medial temporal lobe (MTL) system (Swerdlow & Jicha, 2012). When FCSRT administered in words without immediate recall (wFCSRT) was compared to pFCSRT+IR), free recall was 7.9 points higher and total recall was 4.3 points higher than the respective scores from wFCSRT (Zimmerman et al, 2015). Thus, the scores from these versions are not equivalent and their optimal cut scores will differ.

A limitation of this study is that pFCSRT+IR and LM scores were included with other neuropsychological test scores when diagnoses were assigned, a problem common to many longitudinal studies trying to predict future dementia (Tuokko & Frerichs, 2000). Developing norms in one sample and applying them in an independent sample is ideal. Fortunately, we have such an independent sample (Grober et al, 2010). We applied the cut scores for the conventional and robust norms developed in the present report to an independent sample of 194 adults age 65 or older receiving treatment in a primary care clinic (age =78.3; education =12.5 years). Eligible participants were dementia free at baseline. Twenty-eight incident cases of dementia developed during the three years of follow-up. Sensitivity for detecting the incident dementia cases using the cut score (20) from the conventional normative sample was 21% and was 50% using the cut score (24) from the robust normative sample, with a small decrease in specificity, 96% to 89%.

There are several alternative analytic approaches for using cognitive tests to identify individuals with pre-dementia. If the goal is to identify persons likely to develop dementia over specified time intervals (i.e., 6 months, I year or 2 years), time dependent ROC curves are helpful (Derby et al, 2013). Another approach is to model time to dementia onset in an appropriate time to event model such as proportionate hazards models. In addition, the cognitive test could be used as a screen to identify persons who would then be sent for neuroimaging, cerebrospinal fluid or blood based biomarkers. Biomarker evidence of amyloidosis, neuronal injury, or synaptic dysfunction has been incorporated into the revised criteria for MCI due to AD (Albert et al, 2011) and prodromal AD (Dubois et al, 2014). However, some studies suggest that neither imaging nor CSF biomarkers improve prediction of conversion from MCI to AD over sensitive memory and executive function tests (Richard et al, 2013; Gomar et al, 2011).

In conclusion, when using cognitive test norms to characterize normative aging or to identify incident dementia, consideration should be given to the composition of the normative sample and whether cut scores are adjusted for age and education as these factors can dramatically influence the sensitivity of the test in identifying individuals destined to develop AD.

Acknowledgements

Referemces

Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer‚Äôs disease: Recommendations from the National Institute on Aging-Alzheimer‚Äôs Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's and Dementia, 7(3), 270–279.

American Psychiatric Association. (1994). Diagnostic and Statistical Manual of Mental Disorders (4th ed.). Washington, D.C.: American Psychiatric Association Press.

Brookmeyer R (2007). Forecasting the global prevalence and burden of Alzheimer's Disease. Alzheimer's & Dementia, 3, S168.

Derby CA, Burns LC, Wang C, Katz MJ, Zimmerman ME, L'Italien G, et al. (2013). Screening for predementia AD: Time-dependent operating characteristics of episodic memory tests. Neurology, 80(14), 1307–1314. [PubMed: 23468542]

De Santi S, Pirraglia E, Barr W, Babb J, Williams S, Rogers K, et al. (2008). Robust and conventional neuropsychological norms: Diagnosis and prediction of age-related cognitive decline. Neuropsychology, 22(4), 469–484. [PubMed: 18590359]

Donohue MC, Sperling RA, Salmon DP, & et al. (2014). The Preclinical Alzheimer Cognitive composite: Measuring amyloid-related decline. JAMA Neurology, 71(8), 961–970. [PubMed: 24886908]

Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. (2014). Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. The Lancet Neurology, 13(6), 614–629. [PubMed: 24849862]

Gomar JJ, Bobes-Bascaran MT, Conejero-Goldberg C, Davies P, Goldberg TE, & Alzheimer's Disease Neuroimaging Initiative, f. (2011). Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative. Archives of General Psychiatry, 68(9), 961–969. [PubMed: 21893661]

Grober E, Buschke H, Crystal H, Bang S, & Dresner R (1988). Screening for dementia by memory testing. Neurology, 38(6), 900–903. [PubMed: 3368071]

Grober E, Lipton RB, Hall C, & Crystal H (2000). Memory Impairment on Free and Cued Selective Reminding Predicts Dementia. Neurology, 54, 827–832. [PubMed: 10690971]

Grober E, Sanders AE, Hall C, & Lipton RB (2010). Free and cued selective reminding identifies very mild dementia in primary care. Alzheimer Dis Assoc Disord, 24(3), 284–290. [PubMed: 20683186]

Grober E, Ehrlich A, Troche Y, Hahn S, & Lipton RB (2014). Screening Older Latinos for Dementia in the Primary Care Setting. Journal of the International Neuropsychological Society, 20, 1–8. [PubMed: 24365485]

Hebert LE, Weuve J, Scherr PA, & Evans DA (2013). Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. Neurology, 80(19), 1778–1783. [PubMed: 23390181]

Holtzer R, Goldin Y, Zimmerman M, Katz M, Buschke H, & Lipton RB (2008). Robust norms for selected neuropsychological tests in older adults. Archives of Clinical Neuropsychology, 23, 531–541. [PubMed: 18572380]

Ivnik RJ, Smith GE, & Lucas JA (1997). Free and cued selective reminding test: MOANS norms. Journal of Clinical and Experimental Neuropsychology, 19, 676–691. [PubMed: 9408798]

Katz MJ, Lipton RB, Hall CB, Zimmerman ME, Sanders AE, Verghese J, et al. (2012). Age-specific and sex-specific prevalence and incidence of mild cognitive impairment, dementia, and Alzheimer dementia in blacks and whites: a report from the Einstein Aging Study. Alzheimer Dis Assoc Disord, 26(4), 335–343. [PubMed: 22156756]

Kaduszkiewicz H, Eisele M, Wiese B, Prokein J, Luppa M, Luck T, … Group, D. i. P. C. P. S. (2014). Prognosis of Mild Cognitive Impairment in General Practice: Results of the German AgeCoDe Study. The Annals of Family Medicine, 12(2), 158–165. doi: 10.1370/afm.1596 [PubMed: 24615312]

McNemar Q (1947). Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika, 12 (2) 153–157. [PubMed: 20254758]

Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, & Kokmen E (1999). Mild cognitive impairment: Clinical characterization and outcome. Archives of Neurology, 56(3), 303–308. [PubMed: 10190820]

Richard E, Schmand BA, Eikelenboom P, & Van Gool WA (2013). MRI and cerebrospinal fluid biomarkers for predicting progression to Alzheimer's disease in patients with mild cognitive impairment: a diagnostic accuracy study. BMJ Open, 3(6), 2012–002541.

Ritchie LJ, Frerichs RJ, & Tuokko H (2007). Effective Normative Samples For the Detection Of Cognitive Impairment in Older Adults. The Clinical Neuropsychologist, 21(6), 863–874. [PubMed: 17853155]

Sliwinski M, Lipton RB, Buschke H, & Stewart WF (1996). The effect of preclinical dementia on estimates of normal cognitive function in aging. Journal of Gerontology; Psychological Sciences, 51B, 217–225.

Sliwinski M, Buschke H, Stewart WF, Masur D, & Lipton RB (1997). The effect of dementia risk factors on comparative and diagnostic selective reminding norms. J Int Neuropsychol Soc, 3(4), 317–326. [PubMed: 9260441]

Wagner M, Wolf S, Reischies FM, Daerr M, Wolfsgruber S, Jessen F, et al. (2012). Biomarker validation of a cued recall memory deficit in prodromal Alzheimer disease. Neurology, 78(6), 379–386. [PubMed: 22238414]
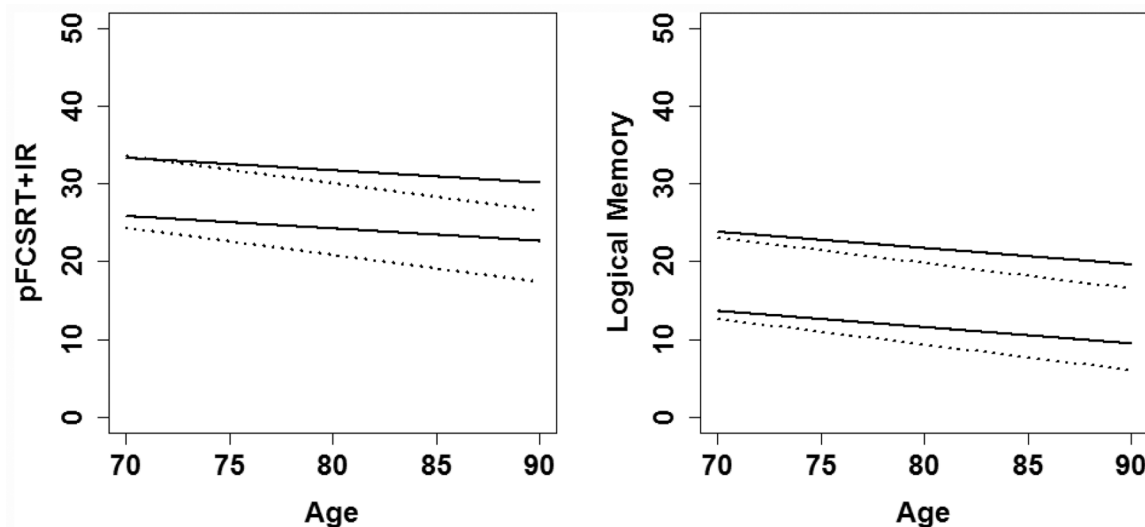
Wechsler D (1997). Wechsler Memory Scale-Revised (1987).

Figure 1.
Estimated mean and 1.5 SD limits for FR and LM as a function of age in robust and conventional samples for subjects with >=12 years of education.
Abbreviations:
pFCSRT+IR: Picture version of the Free and Cued Selective Reminding Test with Immediate Recall
LM: Logical Memory Subtest of the WAIS-R Immediate Recall
Note: The estimated mean and 1.5 SD below the mean, the cut-score, derived from the conventional (dotted line) and robust (solid line) normative samples are shown. Adjusted scores for subjects with <12 years of education are not shown.

Table 1.

Baseline demographic and cognitive characteristics in a conventional normative sample, a robust normative sample and a sample that develops incident dementia within 5 years.

	Conventional normal (n=418)	Robust normal (n=340)	Incident Dementia (n=78)	Incident D vs. Robust Normals
Age, mean (SD), years	78.2 (4.7)	77.4 (4.5)	81.6 (4.4)	p<0.0001
Age Range, years	70.2–89.8	70.2–89.7	72.4–89.8	
Gender				
Female, %	59.3	57.9	65.4	p=0.23
Male,%	40.7	42.1	34.6	
Education (in years)				
<12, %	21.0	18.8	30.8	p=0.02
12, %	79.0	81.2	69.2	
Race				
White,%	67.5	67.9	65.4	P=0.19
Black,%	28.0	26.8	33.3	
Other,%	4.5	5.3	1.3	
Memory tests, mean (SD)				
pFCSRT+IR	30.2 (6.4)	31.9 (5.1)	22.3 (6.4)	p<0.0001
70–79	31.2 (6.2)	32.3 (5.1)		
80–89	28.4 (6.3)	30.8 (4.7)		
LM	19.8 (7.5)	21.1 (7.2)	13.4 (5.7)	p<0.0001
70–79	21.0 (7.3)	21.8 (7.0)		
80–89	17.5 (7.3)	19.4 (7.3)		

pFCSRT+IR: Picture version of the Free and Cued Selective Reminding Test with Immediate Recall

LM: Logical Memory Subtest of the WAIS-R Immediate Recall

Note:

The conventional normative sample is divided into two subgroups, those who develop dementia within 5 years (incident dementia) and those who remain dementia free for at least 5 years. Compared with the robust normal group, incident dementia subjects were older (p<0.0001) and performed worse on both pFCSRT+IR (p<0.0001) and LM tests (p<0.0001) at baseline.

Table 2.

Linear regressions used to compute age and education adjusted cut-scores for pFCSRT+IR and LM in the conventional and robust samples.

Parameter	Conventional			Robust		
	Coefficient	SE	p-value	Coefficient	SE	p-value
	pFCSRT+IR					
Constant	29.08	0.66	<.0001	30.31	0.63	<.0001
Age	−0.35	0.06	<.0001	−0.16	0.06	0.0097
Educ 12	0.65	0.75	0.3853	1.45	0.70	0.0394
	F(2,415)=16.95, p(Pr>F)<0.0001			F(2,337)=6.49, p(Pr>F)<0.0017		
	Multiple R^2 = 0.076			Multiple R^2 = 0.037		
	Residual MSE = 6.146			Residual MSE = 4.992		
	LM					
Constant	15.53	0.75	<.0001	16.53	0.85	<.0001
Age	−0.33	0.07	<.0001	−0.21	0.08	0.0114
Educ 12	4.61	0.86	<.0001	4.99	0.95	<.0001
	F(2,415)=29.99, p(Pr>F)<0.0001			F(2,337)=19.41, p(Pr>F)<0.0001		
	Multiple R^2 = 0.126			Multiple R^2 = 0.103		
	Residual MSE = 6.999			Residual MSE =6.792		

pFCSRT+IR: Picture version of the Free and Cued Selective Reminding Test with Immediate Recall

LM: Logical Memory Subtest of the WAIS-R Immediate Recall

Table 3.

Effects of Norming Strategy and Age and Education-Adjustment on the Sensitivity and Specificity of the pFCSRT+IR and LM for Future Dementia.

Conventional Normative Sample pFCSRT+IR (mean=30.2, SD=6.4, range=9–45, cut-score= 20.6) LM (mean=19.8, SD=7.5, range=1–39, cut-score=8.6)			
	Incident cases identified out of 78	Sensitivity	Specificity
Unadjusted cut-sore pFCSRT+IR	27	34.6	97.9
Adjusted cut-score pFCSRT+IR	25	32.0	97.1
Unadjusted cut-score LM	15	19.2	95.9
Adjusted cut-score LM	14	18.0	96.5
Robust Normative Sample pFCSRT+IR (mean=31.9, SD=5.1, range=14–45, cut score=24.3) LM (mean=21.1, SD=7.2, range=1–39, cut score=10.3)			
	Incident cases identified out of 78	Sensitivity	Specificity
Unadjusted cut-sore pFCSRT+IR	45	57.7	92.4
Adjusted cut-score pFCSRT+IR	40	51.3	92.7
Unadjusted cut-score LM	25	32.0	92.7
Adjusted cut-score LM	21	26.9	92.4

pFCSRT+IR: Picture version of the Free and Cued Selective Reminding Test with Immediate Recall

LM: Logical Memory Subtest of the WAIS-R Immediate Recall

Table 4.

The contingency table comparing pFCSRT+IR and LM in detecting incident dementia from baseline data.

pFCSRT+IR		LM Incident Dementia		
		No	Yes	Total
Incident	No	25	8	33
Dementia	Yes	28	17	45
	Total	53	25	78

pFCSRT+IR: Picture version of the Free and Cued Selective Reminding Test with Immediate Recall

LM: Logical Memory Subtest of the WAIS-R Immediate Recall